Workshop on
Air Quality and Atmospheric Composition Metadata
Morrison Hotel, Dublin, Ireland
5-7 September 2012

Convenors: Martin Schultz, Leonor Tarrason
Local organizer: Colin O'Dowd

Sponsored by:

macc
Monitoring atmospheric
composition & climate

SEVENTH FRAMEWORK
PROGRAMME

JÜLICH
FORSCHUNGSZENTRUM

epa
Environmental Protection Agency
An Ghníomhaireacht um Chaomhnú Comhshaoil

NILU

EEA

Historical Background

1995: the **Dublin core** metadata initiative laid the foundation for standardizing discovery metadata. It laid out a vocabulary of fifteen properties for use in resource description. (but: Dublin, Ohio)

2005: GEO was established with a mandate to build a Global Earth Observation System of Systems (GEOSS)

2007: the **INSPIRE directive** came into force with the goal to create a European Union (EU) spatial data infrastructure

2009: **MACC** project starts as GMES pre-operational service; **EEA** gets mandate as GMES in-situ data coordination center

2011: the **GEO AQ CoP** had ist first physical meeting on Solta, Croatia

Networking Air Quality Observations and Models:
From Virtual to Real

Aug 23-25, 2011, Stomorska, Šolta, Croatia

Main focus: technical realization of interoperable air quality data network
Discussion topics:
- Interoperability standards
- WCS server developments and requirements
- Impediments for realizing a global air quality data network

"What few things must be the **SAME**,
so that everything else can be **DIFFERENT**?"

Workshop

„Metadata for Air Quality and Atmospheric Composition"

Morrison hotel, Lower Ormond Quay, Dublin

5 – 7 September, 2012

One of main impediments identified on Solta was incomplete standardisation and diverging knowledge about metadata, realizing that good metadata are key to achieve interoperability.

Therefore, the main goal for Dublin 2012 is to develop **good practices** for metadata!

Your Background

	OBS	EMI	SAT	MOD	OTHER
single	12	0	2	2	2
multiple	15	6	7	11	2

familiar with…

GEO CoP	INSPIRE	MACC	netCDF	GEIA	GISC
11	14	22	17	6	5

GCMD	ESIP	Datafed	Eurogeoss-Broker
4	5	5	3

Why care about metadata?

What you want are the data (i.e. chocolate)

How do you find out if this is what you want?

1. Buy it, taste it

2. Read the package information

3. Evaluate metadata record and order from online shop

What do you need to know before you buy?

Brand name (dataset title)

Type of chocolate (dataset type)
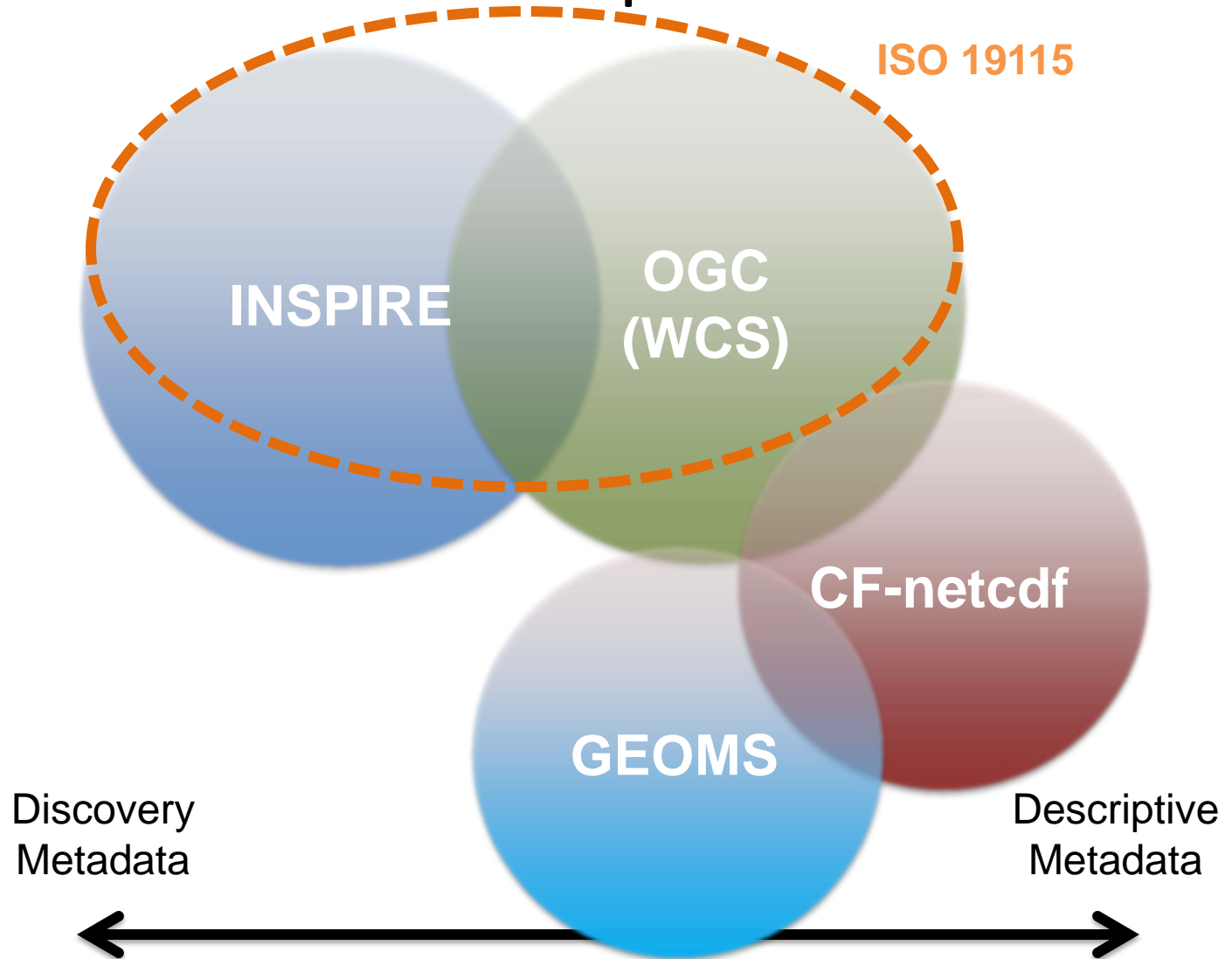
Size of chocolate (geospatial extent)

Ingredients (dataset quality)

Best before (dataset history)

The **FAUL** concept of interoperability

FIND
ACCESS
UNDERSTAND
LINK

without having to program anything as a user

FAUL (as a German word) means „lazy"

So, what is metadata really?

Suggested definition:

All information required to be **FAUL**

What is needed to discover AQ/AC data?

Option 1:
Unstructured search
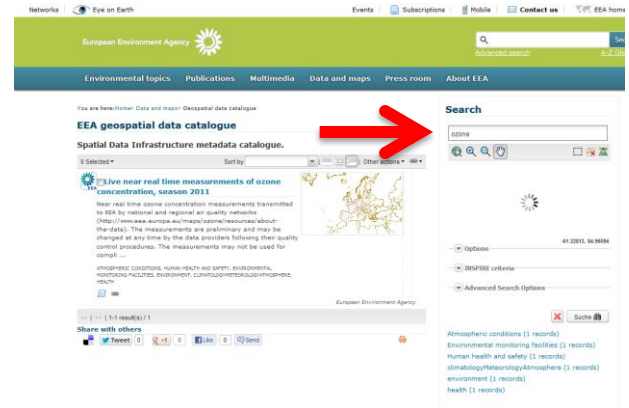
Option 2:
Faceted search
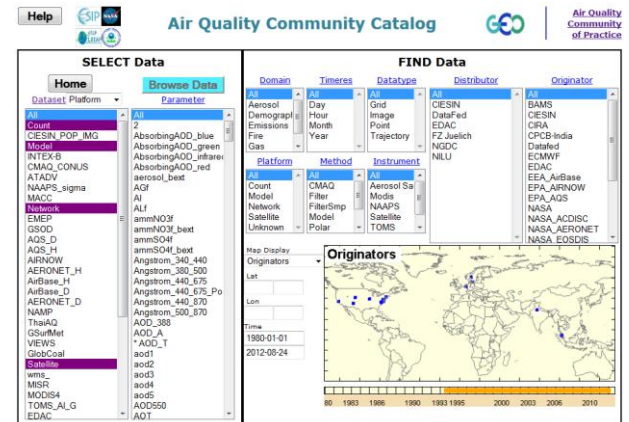
What is needed to access AQ/AC data?

A

Type of service (ftp, wcs, …)
Service URL
Service provider
Access/Use constraints

But before, you want to know about:
Suitability of data (for example resolution)
Data format
Data volume (?)
„Freshness" of product

What is needed to understand AQ/AC data?

U

Where do the data originate from?

How were data produced?

What is the data quality?

What changes have been applied to the data?

etc.

What is needed to link AQ/AC data?

Some requirements:

Common (interchangeable) formats

„Intelligent" dataset organisation

Value-added products

Interchangeable units of measure

Workshop Objectives

- Identify essential metadata elements needed for dataset and service discovery and use

- Map this information onto existing standards
  (and check if terminology can be translated between standards)

- Advance common understanding of interoperability
  (connections among networks, servers, people)

- **Formulate best practices for metadata generation and use**

Expected Outcomes

- A set of „good practice" recommendations for metadata generation and use across communities

- Concrete actions to communicate these findings
    - MACC Report
    - Review of existing proposals for e-reporting/data flagging from GISC
    - Publication in IEEE-JSTAR
    - Identify how CoP AQ can contribute to the GEO Workplan
    - Link to INSPIRE
    - Implementation in selected networks

Workshop structure

Day 1 (Wednesday): Presentations

   INTRODUCTION - *Main frameworks dealing with AQ metadata*

   BLOCK 1 - *Metadata needs from different user communities*

   BLOCK 2 - *Metadata standards in practice: what do we get?*

Day 2 (Thursday): Working Groups

   SESSION 1: Map current practices and identify limitations

   SESSION 2: Propose best practices and uses

Day 3 (Friday): Synthesis of best practices & Communication

Participate
and
enjoy the meeting!

Working Group Discussions

Main Objective:

Development of good practices for metadata characterisation

Agenda WG discussion

SESSION 1: THURSDAY MORNING

09:00 -12:30 **Objective**:   Map current practices and uses
                                     Identify main limitations in current metadata standards
                   4 groups: Observed data
                                     Modelled data
                                     Satellite Data
                                     Emission data

12:30 - 14:00 Lunch Break

14:00 - 16:00 Report from working groups to plenary -
                   Life demonstration of metadata ingestion into GI-Cat
                   Plenary agreement on **main limitations** in current standards

SESSION 2: THURSDAY AFTERNOON

16:30 – 18:00 **Objective**:   Propose solutions - **best practices and uses**
                   4 groups: Discovery
                                       External Properties
                                       Ownership and Provenance
                                       Data Description
              Report to plenary on Friday morning

Discovery of the data

1. **How can we identify data sets?** (e.g. master list(s) of species/compounds, instruments, etc.)
2. **What does „Data model" mean wrt AQ data?** Please map different terminologies
    - „observation" – „process" (INSPIRE terminology)
    - „dataset", „series", „service" (ISO terminology)
    - data types (cf. INSPIRE AC-MF)
3. **Metadata for unstructured discovery (search bar)**
    - Can you provide standardized keywords, lists for AQ data? (GEMET, GCMD, …)
4. **Metadata „facets" for structured discovery (e.g. core.uFind)**
    - review of current facets
    - mapping of facets terminology onto existing standards
    - use of vocabulary lists

External Properties

1. **How can we describe external data set properties?**
   Provide recommendations on how to provide station description; instrument description;calibration methods; etc.)

2. *External" (descriptive) metadata (e.g. station description)*
   - how to describe what can be found?
   - how to link to the data description?

-

Ownership and Provenance

1. How can we describe data set ownership and provenance?

- Who "created" the data set?
- Who is responsible for which aspect of the data and metadata?)

2. How to harmonise Roles": *the process of data and metadata provision*
- „PI", „submitter", „provider" (GEOMS)
- „originator", „provider", „distributor" (INSPIRE, ISO)
- what defines identity of data | series | service | metadata? (TH doc)

3. How to best describe data set provenance (history)
- how to describe what has been done to the data?
- how to uniquely identify a dataset (and its ancestors)?

Data description

1. How can we track data set version changes and data modifications? (updated calibration records; aggregated or filtered data; interpolated data; reformatted data; etc.) –

2. What do we need to describe the data set contents? (measurement vs.model; analysis vs. forecast; resolution; continuous series vs. campaign data; multi-species set; etc.)

3. How do we describe the data set quality? how to include data quality information?

4. XML best practices

- structure (sections) of XML file

- description of spatial and temporal extent (also vertical)

- how to link to vocabulary servers?

- how to set-up metadata servers?

- when to update metadata records?

- how to publish metadata records (and where)?

Synthesis

- Identify specific suggestions to enhance standards (ISO, INSPIRE, CF, …)

- Identify options to improve „metadata flow"

- Identify how to improve links among existing catalogues, services, data sets, people

PROVIDE RECOMMENDATIONS FOR BEST PRACTICES

SESSION 1
THURSDAY 6th 09:00-12:30

Objective: Map current practices and uses
Identify main limitations in current metadata standards

1. Observed data   (PRINTWORKS 1)
2. Modelled data   (PRINTWORKS 2)
3. Satellite Data   (GEORGIAN BOARDROOM)
4. Emission data   (STRAND STREET)

We meet at coffee break at 10:30-11:00 !

SESSION 2

THURSDAY 6th 16:30-18:00

Meets after coffee break from 16:00-16:30 !

Objective: Propose solutions – Identify best practices and uses

1. Discovery   (PRINTWORKS 1)
2. External Properties   (PRINTWORKS 2)
3. Ownership and Provenance   (STRAND STREET)
4. Data Description   (GEORGIAN BOARDROOM)

Reports back to plenary on Friday morning